YEARBOOK OF PHYSICAL ANTHROPOLOGY 42:127–151 (1999)

Quantitative Trait Linkage Mapping in Anthropology

J. ROGERS,[1,2] M.C. MAHANEY,[1] L. ALMASY,[1] A.G. COMUZZIE,[1]
AND J. BLANGERO[1]
[1]*Department of Genetics, Southwest Foundation for Biomedical Research, San Antonio, Texas 78245*
[2]*Southwest Regional Primate Research Center, P.O. Box 760549, San Antonio, Texas 78245*

KEY WORDS quantitative trait; gene mapping; genetics; genomics

ABSTRACT Recent years have seen rapid progress in several areas of both biomedical and anthropological genetics. While genetic analyses have come to play a significant role in biological anthropology, there has been little use of modern methods for linkage mapping of quantitative trait loci (QTLs). It is now feasible to design research studies to investigate the quantitative genetics of complex phenotypes that are of primary importance to traditional questions in biological anthropology. Complex traits such as functionally significant morphological features, physiological characteristics or aspects of behavior can be examined to estimate the influence of genetic variation on within-species phenotypic variation. In addition, new methods for mapping quantitative trait loci provide opportunities to identify the regions within chromosomes that contain the functional genes of interest. This review summarizes molecular genetic and statistical genetic approaches to QTL mapping, and presents examples of how this approach can expand the scope of anthropological genetics to include mapping and identifying individual genes that influence complex phenotypic traits relevant to fundamental questions in biological anthropology. Yrbk Phys Anthropol 42:127–151, 1999.

TABLE OF CONTENTS

GLOSSARY

Complex phenotype: Any phenotype whose expression is influenced by multiple genes, or by one or more genes and one or more environmental factors. Complex traits can be quantitative or discrete, but most discrete complex traits are the result of underlying quantitative liability functions.

Heritability (broad sense): Proportion of total phenotypic variance that is attributable to genetic variation.

Heritability (narrow sense): Proportion of total phenotypic variance that is attributable to additive genetic variation.

Identity by descent (IBD): Two alleles at a locus are identical by descent when they have been inherited without new mutation from a single common ancestor.

Identity by state (IBS) : Two alleles at a locus are identical by state when they are of the same form, whether or not they are derived from the same ancestral allele.

λ_R: Relative risk of having a disease given an affected relative of degree R. It is given by the ratio of the absolute risk to the relative and the general population prevalence.

Linkage disequilibrium: Nonrandom association within a population of alleles at 2 or more linked loci. Linkage disequilibrium decays with increasing genetic (recombination) distance between loci.

LOD score: Logarithm of the odds. The ratio of two likelihoods (likelihood of a specific data set given linkage between two loci and likelihood of the same data given no linkage) that is used as a measure of the statistical support for the hypothesis that two loci are linked.

Microsatellite: A short segment of DNA that contains tandem repeats 2–5 base pairs in length, and shows polymorphism in the number of repeats. These loci are also called simple sequence repeats (SSR) or short tandem repeats (STR).

Monogenic: A phenotypic trait is monogenic if that trait is influenced primarily or entirely by only one genetic locus.

Oligogenic: A phenotypic trait is oligogenic if it is influenced by a few loci of significant, individually detectable effects.

Penetrance: The probability that an individual carrying a particular genotype will express the trait of interest. A trait with low penetrance will be expressed by only a small proportion of individuals in a population who carry the susceptible genotype. A phenotype with high penetrance is expressed by a larger proportion of individuals carrying the susceptible genotype, regardless of the overall prevalence of the trait.

Polygenic: A phenotype is polygenic if it is influenced by many genes of relatively small individual effects, such that the influence of any single locus is very difficult or impossible to detect on its own. In genetic theory, polygenic traits are often modeled as if they are influenced by an infinite number of loci with equal effects.

Prevalence: The overall proportion of individuals in a population that exhibit a given phenotype or genotype.

Quantitative trait locus: Any locus that influences variation in a complex phenotype.

Single nucleotide polymorphism (SNP): Any single DNA base pair change that is polymorphic within a given population.

Analyses of the amount and nature of genetic variation in populations of humans and nonhuman primates, both extant and extinct, have come to play an increasingly significant role in biological anthropology. It is now taken for granted that information on DNA sequences is critical to studies of species-level phylogenetics, the affinities and origins of living human populations, and other traditional anthropological questions. The recent confluence of methodological developments in quantitative and statistical genetics with those of molecular genomics, originally motivated to find genes responsible for complex human disorders, has made it possible to further expand the role of genetic analysis in this discipline. In particular, the Human Genome Project, related

research efforts involving laboratory animal model species, and other biomedical initiatives have yielded methodologies and resources that are likely to produce important advances in biological anthropology. If investigators choose to master and apply these new techniques, the mapping and identification of specific genes that influence inter-individual variation in phenotypes of interest to biological anthropology — e.g., body size, rates of growth and maturation, bone shape, dental morphology, reproductive physiology or behavior — are now feasible.

The goal of this paper is to present an overview of the methods used to localize, within the human and nonhuman primate genomes individual genes that influence complex phenotypic traits. These methods include both molecular genetic analysis of DNA sequence variation and statistical genetic analysis of the relationships between DNA polymorphisms and phenotypic variation. We also discuss examples of how these methods might benefit traditional research questions in biological anthropology. This is not a typical literature review, in that little of this type of research has yet been completed and published. As a result, this is less a review of past research progress than a description of currently available tools and potential research goals. If readers find this to be a useful, although brief, overview of an established and rapidly developing subfield in biomedical genetics and its possible relevance to the future of anthropological genetics, then we will have achieved our goal.

THE NATURE OF QUANTITATIVE TRAIT LOCI

Complex phenotypes

The approaches described in this review have been developed to detect and localize genes influencing phenotypic variation in complex traits that are either disease entities themselves or risk factors for diseases. A complex phenotype is any measurable characteristic of an organism that is influenced by both genes and environmental variables to an extent that simple Mendelian patterns of inheritance are not readily discernable. (Note that some authors describe complex phenotypes as those exhibiting "non-Mendelian" inheritance.) This description will apply to most of the traits whose patterns of within-species phenotypic variation are of interest to biological anthropologists.

Although commonly used to describe traits that vary and are measured on a continuous scale, such as stature, body weight, red blood cell number, etc., the term "complex" refers to the sources of inter-individual variation for a specific trait, rather than whether a trait is continuous or discontinuous in its distribution within a population. Multiple genes, multiple environmental factors, or some combination thereof influence patterns of phenotypic variation in complex traits. In fact, ostensibly all-or-none disease states, such as myocardial infarction, diabetes, and schizophrenia, all may be classified as "complex" traits because variation in their severity, age at onset, and symptoms cannot be attributed to a single gene or environmental factor. While genetic syndromes and diseases attributed to major mutations are often cited as exceptions, single genes are unlikely to be responsible for all the phenotypic variation observed among the individual carriers. Indeed, even in relatively "simple" genetic disorders such as Down's syndrome, cystic fibrosis, or phenylketonuria that are frequently cited as examples of simply inherited and/or monogenic states, phenotypic heterogeneity due to factors other than the primary mutation itself is common (see, e.g., Raskin et al., 1999; Kayaalp et al., 1997; Schellenberg et al., 1992).

Biological anthropologists have concentrated predominantly, but not exclusively, on normal phenotypic variation in features of human and nonhuman primate anatomy, morphology, physiology, growth, and behavior. Most of the phenotypic variation within or between species in traits of interest to biological anthropologists is likely to be attributable in part to the effects of genes at multiple chromosomal loci. In addition, the expression of genes at these loci will probably be influenced by one or more environmental factors to produce the observed patterns of variation in these phenotypes. Because most of these traits are measured on a continuous or quantitative scale, they are commonly referred to as quantitative traits and have been subject to quantitative genetic analysis. Loci harboring genes that contribute to the patterns of variation in such phenotypes have been given the appel-

lation "quantitative trait loci" or QTLs (Geldermann, 1975). Any one QTL may account for only a small proportion of the phenotypic variance observed in a trait, with other QTLs and environmental factors accounting for the remainder.

The QTL concept has been expanded to include loci influencing complex traits exhibiting discontinuous variation, such as schizophrenia or diabetes. The risk of developing these diseases is generally thought to be a function of an underlying continuous quantitative liability function. Discrete morphological traits such as supernumerary teeth or wormian bones would be examples of such phenotypes from the biological anthropology literature.

Quantitative genetics

Estimation of the relative proportions of the total phenotypic variance in a complex trait that is attributable to the effects of genes is fundamental to locating and identifying individual QTLs. In accordance with classical quantitative genetics theory (see Falconer, 1989; Lynch and Walsh, 1998), the total variance in a trait, σ_P^2 may be decomposed into its genetic and environmental components such that $\sigma_P^2 = \sigma_G^2 + \sigma_E^2$, where σ_G^2 is the genetic component to the variance and σ_E^2 is the environmental (i.e., nongenetic) component. Each of these components may be decomposed further. For example, σ_G^2 can be decomposed into components representing the variance attributable to additive genetic effects (σ_A^2), dominance (σ_D^2), and/or epistasis (σ_I^2); while σ_E^2 can be decomposed into components attributable to measured environmental factors and random, unmeasured nongenetic factors. The proportion of phenotypic variance in a trait that is due to genetic differences among individuals is generally referred to as the heritability of the trait, and symbolized as h^2. The proportion of phenotypic variation attributable to *all* genetic effects (e.g., additivity, dominance, epistasis) is referred to as the "broad sense" heritability and obtained as σ_G^2/σ_P^2 while the heritability in the "narrow sense" refers to the proportion attributable to the *additive* genetic variance alone, i.e., σ_A^2/σ_P^2 (Falconer, 1989; Lynch and Walsh, 1998). Unless otherwise specified, we shall be using the term "heritability" in this latter, narrow sense. Given the additive nature of the components of the phenotypic variance, it can be seen that the heritability of a trait will be influenced by the magnitude of the underlying genetic variance and the amount of environmentally induced variation. Consequently, any estimate of heritability must be specific to a given population and environmental context.

Once a trait is known to be heritable (i.e., some proportion of phenotypic variance is due to the effects of genetic differences among individuals) investigators may want to determine the chromosomal locations of the individual genes that account for the genetic effects. Theoretically, *any* locus that influences this variability may appropriately be classified as a QTL. However, the power of most available analytical approaches to detect a QTL is a complex function of the effect size (or heritability) of the QTL itself, the study design (i.e., data obtained from sib-pairs, sibships, nuclear families, or extended pedigrees), sample size, and the characteristics of the genetic data (i.e., number and heterozygosity of molecular genetic markers). Consequently, detection and localization of genes with very small effects may not always be feasible. In practice, QTL detection and localization will be more successful for traits that are monogenic (influenced primarily or entirely by a single locus) or oligogenic (influenced by a few genes, each with substantial effects) than for traits whose phenotypic variation is attributable to dozens of genes with equal and individually small effects. By means of the analytical methods discussed below, it is now reasonable to expect that, with appropriate sample sizes and study designs, individual genes accounting for 10–15% of variation in a trait can be localized to specific chromosomal regions (Williams and Blangero, 1999a,b; Williams et al., 1997).

Other experimental strategies that do not use human or nonhuman primate subjects can detect effect sizes smaller than the 10–15% level discussed above. For example, studies of inbred strains of mice have detected loci responsible for less than 4% of the phenotypic variance in traits related to growth and development (Cheverud et al.,

1996). Inbred animal models provide the researchers with a number of advantages when conducting searches for QTLs. These include, but are not limited to, the ability to (a) eliminate variation at loci other than the ones being studied and (b) maximize control of environmental sources of phenotypic variation. Doing so decreases the remaining, or residual, variance to be explained and increases the relative proportion of total phenotypic variance that is attributable to the effects of genes of interest.

Analogous effects can sometimes be achieved in analyses of data from non-inbred human or nonhuman primate populations. By accounting for the variance in a phenotype that is attributable to environmental factors, and even that which is due to other genes, one can perform sequential analyses that progressively dissect the genetic components of a phenotype. For example, assume that additive genes account for 10% of the *total* variation in a phenotype, *X*. On initial consideration, this would seem to indicate that we are at the limit of detection for a QTL for *X* and, in fact, probably would be successful only if a single QTL accounted for *all* the additive genetic variance in *X*. Now, assume that 60% of the total phenotypic variation in *X* is attributable to nongenetic factors (covariates) such as age, sex, and/or an environmental variable such as diet. If that is the case, the proportion of the *residual* phenotypic variance in *X* attributable the additive effects of genes (which we will refer to as h_R^2) would be obtained as the quotient of the original heritability, h^2, and the proportion of the phenotypic variance not accounted for by those covariates. That is

$$h_R^2 = h^2/(1 - 0.60) = 0.10/0.40 = 0.25.$$

Note that the actual contribution of genes to the *total* phenotypic variance in *X* has not changed, but accounting for the effects of covariates on *X* improves the genetic signal-to-noise ratio and, in this case, more than doubles the heritability estimate. Identifying additional salient covariates can improve the residual heritability even more.

As described above, power to detect a QTL for common, continuously varying traits, is a function of, among other factors, the proportion of the residual phenotypic variance attributable to the QTL. Because it is the residual heritability — or rather the genetic variance underlying it — that is decomposed further in a such an analysis, accounting for covariate effects can improve the likelihood that one will find QTLs with relatively small effects on the total variance. Further, once a valid QTL has been detected, it also can be incorporated into subsequent analyses to redefine further the residual phenotypic variance. This allows for sequential searches for QTLs with smaller and smaller effects (Almasy and Blangero, 1998).

The brief discussion above establishes no absolute lower limit on detectable effect sizes for QTLs. However, the ultimate objectives and applications of a research project will determine whether the investigator wishes to pursue QTLs of very small effect. For example, if the objective is to develop a complete understanding of the genetic contributions to variability in a metabolic pathway (i.e., to identify all the loci that influence any aspect of that pathway), then localizing and identifying as many QTLs as possible, regardless of their effect sizes, may be justified. However, this will require a large investment in terms of sample sizes and analytic effort. By contrast, if the objective is to identify QTLs whose effects are substantial, and important for the determination of inter-individual variation in growth, development, health, adaptation, etc., in a population, then efforts to identify a locus responsible for 1% or less of the variance may not be warranted.

METHODS FOR MAPPING QTLs

Genetic markers for gene mapping

Once a given phenotype is shown to be heritable, investigators will next want to locate individual quantitative trait loci on chromosomes. This is done through quantitative trait linkage analysis. In order to perform linkage analyses, it is first necessary to generate an appropriate chromosomal map of genetic markers in the pedigrees to be studied. For a robust analysis, one should have genotype data available for a series of highly informative, i.e., highly heterozygous, genetic markers spaced evenly across the chromosomes of every individual

to be used in the linkage study. Marker-to-marker linkage analysis should also be done to estimate the recombinational distances between loci in the map.

Microsatellites, or simple sequence repeat loci, have become the dominant type of genetic marker for linkage analyses. These are highly polymorphic loci that consist of tandem repeats of 2–5 base pairs (bp) (Weber and May, 1989; Bruford and Wayne, 1993). The repeat units are highly susceptible to mutations that increase or decrease the number of repeats. As a result, a population will accumulate a large number of alleles at a microsatellite locus, where the recognized alleles differ in the number of repeat units. The length polymorphism that results is readily analyzed using polymerase chain reaction (PCR) methods and gel electrophoresis. The human genome map now includes more than 8,000 microsatellites, and similar DNA polymorphisms have been described for mice, rats, cattle, and other mammals (for reviews, see the *ILAR Journal* 39:47–256,1998). A smaller number of polymorphic microsatellites have been published for chimpanzees (Deka et al., 1994; Morin et al., 1994), rhesus macaques (Morin et al., 1997), baboons (Rogers et al., 1995, Perelygin et al., 1996, Morin et al., 1998) and other primates (e.g., Witte and Rogers, 1999).

The statistical power of a genetic linkage analysis will depend, in part, on the heterozygosity of the genetic markers used. Markers of low heterozygosity are less useful in determining which alleles are identical by descent among a set of individuals, and therefore statistical power is reduced (see below). The average spacing among polymorphic markers is also a factor in determining statistical power, since functional genes that occur in large gaps within a map are less likely to be detected than genes more closely linked to a genotyped marker. The generally accepted density of information for initial linkage analyses is a linkage map of markers with average heterozygosity about 0.70 or higher, and average spacing less than or equal to 10 centiMorgans (cM). This rule of thumb is based on simulation studies, and to some extent empirical experience.

Recently, substantial attention has turned to another type of human genetic marker. Single nucleotide polymorphisms (SNPs) are single DNA bases that are polymorphic within the human genome. Base pair changes in intergenic regions, or within introns of functional genes, are reasonably common. About one nucleotide out of 300–1,000 is estimated to be variable in most human populations (Li and Sadler, 1991; Schafer and Hawkins, 1998). The majority of SNPs are likely to be selectively neutral and, like microsatellites, can serve as genetic markers in linkage analyses. The average heterozygosity of SNPs will be much lower than that of microsatellites, because most SNPs will have just two alleles, and therefore a maximum heterozygosity of 0.50. Consequently, each SNP provides less linkage information in a pedigree study than would a microsatellite at the same chromosomal location. However, the SNPs are apparently much more numerous (Schafer and Hawkins, 1998; Buetow et al., 1999). In addition, new technologies have been developed that allow hundreds of different SNPs to be analyzed very rapidly (Ramsey, 1998; Schafer and Hawkins, 1998). It is anticipated that full genome maps including tens of thousands of SNP markers will be developed to supplement the microsatellite-based linkage map (Pennisi, 1998). The high-throughput genotyping technologies and SNP-based maps will improve linkage analyses, and also permit rapid full-genome association analyses designed to map QTLs through population association in sets of unrelated people (as opposed to linkage analysis in pedigrees). The markers and methods for the genome-wide association approach to QTL mapping are being developed, but the approach has inherent and substantial weaknesses as well as strengths when compared to linkage methods (Pennisi, 1998; Terwilliger and Weiss, 1998).

Penetrance model-based methods for linkage analysis

The genetic determinants of many monogenic diseases have been successfully mapped in extended human pedigrees using classical penetrance-based linkage analysis methods. Classical penetrance model-based methods require detailed knowledge of the underlying genetic model. For a disease

trait, one must specify the prevalence of the trait, its mode of inheritance, the allele frequencies at the presumed disease gene, and the probability of being affected, or penetrance, for each genotype. For quantitative traits, one must specify allele frequencies and the mean trait value for each genotype. For complex phenotypes, which by definition are influenced by multiple genes with penetrance structures or genotypic means that may be dependant on other genes or on unknown environmental factors, specification of the necessary parameters is problematic. Incorrect specification of these parameters reduces the power to detect linkage with the penetrance model-based methods (Greenberg and Hodge, 1989; Greenberg, 1990; MacLean et al., 1993) and can lead to spurious negative findings, including false exclusion of a region containing a QTL (Risch and Giuffra, 1992). To its credit, when a correctly specified model *is* identified and used, the statistical power of the penetrance model-based approach equals or exceeds those of other methods. Further, the results of a penetrance model-based analysis can be used to exclude linkage with a marker locus (e.g., Mahaney et al., 1995). However, given that each correctly specified model describes the penetrance function for only one of perhaps several oligogenes influencing variation in a trait of interest, practitioners may find this an inefficient approach for whole genome scans.

The penetrance-model based linkage methods use marker genotype data directly in their analyses. This is true of two emerging, but not yet widely used, parametric linkage methods. The Markov chain Monte Carlo (MCMC) method (Guo and Thompson, 1992; Heath, 1997) combines segregation and linkage analysis to estimate penetrance-model properties, such as QTL frequencies and genotypic effects as well as QTL position. Although the MCMC method has the flexibility to accommodate complex multilocus genetic models in large pedigrees, it is computationally intensive. Another genotype-based method under development is the use of neural networks that are trained to model phenotype endpoints as linear and nonlinear functions of genotypes (Lucek and Ott, 1997). This method can use numerous phenotypes in a multivariate analysis, easily incorporates potential covariates, and has the ability to consider the entire genome simultaneously. As one might expect, this method is also computationally intensive. Perhaps the most notable current drawback to neural network approaches is a lack of statistical tests for the assessment of significance for their linkage findings.

TABLE 1. Methods of QTL linkage analysis

Linkage method	Reference	Available software[1]
Classical penetrance-based	Ott, 1991	LINKAGE MENDEL PAP SAGE
Haseman-Elston	Haseman and Elston, 1972 Elston et al., 1998	SAGE MAPMAKER/SIBS GAS
Variance components	Almasy and Blangero, 1998 Hopper and Matthews, 1982	SOLAR SEGPATH ACT MIM
Relative pair	Kruglyak and Lander, 1995 Risch, 1990	MAPMAKER/SIBS ASPEX GENEHUNTER GENEHUNTER PLUS SAGE GAS SIMIBD
Markov chain Monte Carlo	Heath, 1997 Guo and Thompson, 1992	LOKI BUGS SIMWALK GAP

[1] Most of these computer programs are available over the internet. For links to these and other programs, visit the list of genetic analysis software maintained by the Statistical Genetics Laboratory at Rockefeller University at http://linkage.rockefeller.edu/soft/list.html.

Penetrance model-free methods for linkage analysis

In response to the need for more efficient linkage analysis methods specifically designed to address complex traits, penetrance model-free methods of analysis are being developed. The following discussion gives an overview of the most widely used complex trait study designs and penetrance model-free linkage methods (see Table 1), as well as their relative effectiveness for different classes of phenotypes.

Penetrance model-free methods are often referred to as "non-parametric" linkage methods, although this is not entirely correct, as these methods involve the estima-

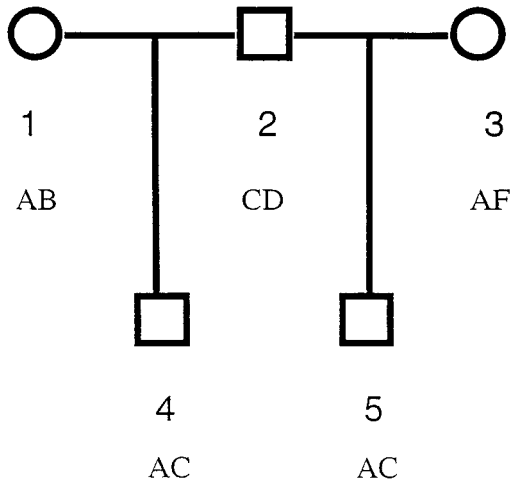

Fig. 1. A simple pedigree illustrating IBD and IBS. Circles represent females, squares represent males, numbers identify individuals in the pedigree, and two letter combinations represent individual genotypes at a polymorphic marker locus.

tion of various parameters other than those having to do with the mode of inheritance. Penetrance model-free methods for linkage analysis are based on the concept of identical by descent (IBD) allele sharing among relatives. Two alleles are considered IBD if they originate from the same ancestral source. In Figure 1, individuals 4 and 5, who are half-siblings, share the allele *C* IBD because both received it from a common ancestor, their father. Although they both have an *A* allele, these alleles are not IBD because they were inherited from the unrelated mothers of the two half-siblings and cannot be traced back to a common ancestor. Allele sharing is also sometimes discussed in terms of identity by state (IBS) in which the ancestral source of the alleles is not considered. Two alleles are IBS if they are of the same form, regardless of their origin. In Figure 1, individuals 4 and 5 share both of their alleles IBS. Additionally, while individuals 1 and 3 can not share any alleles IBD, as they are unrelated, they share the *A* allele IBS. IBD allele sharing provides information about linkage, while IBS allele sharing provides information about population level association between the marker and the trait of interest. However, in population isolates with a modest number of founders IBS allele sharing very closely approximates IBD allele sharing for highly polymorphic markers because each allele is likely to have entered the population through only one founder. We also note that microsatellite alleles that are IBS should not be considered IBD, because their mutation rate is so high that parallel or convergent mutations will be too common.

In the absence of inbreeding, a pair of individuals can share 0, 1, or 2 alleles IBD and the degree of IBD sharing is usually expressed as the proportion of alleles shared (π) which can take the values 0, ½ and 1. The expected IBD sharing over the entire genome ($E(\pi)$) for a relative pair is equal to twice the pair's kinship coefficient (2ϕ). For siblings, $E(\pi)$ is ½, whereas it is ¼ for half siblings and ⅛ for first cousins. If a pedigree is inbred (i.e., there are matings between related individuals), higher degrees of IBD sharing are possible when individuals are homozygous for an allele from a particular ancestor.

Calculation of IBD sharing for a genotyped marker locus is straightforward for nuclear families, but becomes more complicated for extended pedigrees in which the top few generations are often unavailable for genotyping. A number of methods have been proposed to calculate IBD probability matrices (Amos et al., 1990; Whittemore and Halpern, 1994; Curtis and Sham, 1994). One simple and effective approach described by Curtis and Sham (1994) uses maximum likelihood methods to sequentially estimate the IBD probability for each pair of individuals within a pedigree using the posterior probability of genotypes at a completely linked pseudo-marker. For more complex extended pedigrees, a Monte Carlo method can be employed in which missing genotypes are imputed (Almasy and Blangero, 1998). IBD is calculated according to the algorithm of Davis et al. (1996), and a final IBD matrix is estimated based on a weighted distribution produced over many Monte Carlo replicates.

The penetrance model-free linkage methods discussed below each have their advantages and disadvantages. The choice of which to use and in what type of sample depends heavily on the phenotype being analyzed.

Recent methodological studies have suggested that quantitative phenotypic traits measured on a continuous scale are more informative for linkage analysis than dichotomizations of the same trait (Bailey and Almasy, 1995; Duggirala et al., 1997; Korczak and Goldstein, 1997; Wijsman and Amos, 1997). If values of a phenotype of interest are continuously distributed, or that phenotype has quantitative correlates, a linkage analysis of that trait as a continuous scale variable will have more power than an analysis based on categorization of individuals into high and low or affected and unaffected classes.

For traits that are necessarily measured on a discontinuous scale, the choice of sample and analytical method is most dependent on the prevalence of the phenotype. Affected relative methods are often used for such traits, and the statistical power of affected relative pair methods is maximized when population prevalence of the disease/classification is low. Variance component methods are most powerful when disease prevalence approaches 50%. Additionally, the collection of data from extended pedigrees within populations that are randomly ascertained with respect to phenotype offers demonstrably greater power to detect QTLs for common, continuously varying complex traits. However, such a scheme can be rendered useless in searches for genes influencing traits with low prevalence. In such cases, nonrandom ascertainment, i.e., constructing samples using relatives of previously identified individuals who are affected with the trait of interest, can be immensely helpful.

Haseman-Elston methods

Until recently, the most widely used approach for quantitative trait linkage analysis has been the Haseman-Elston method (Haseman and Elston, 1972). This method is based on regressing the squared difference in trait values for pairs of siblings on the proportion of alleles shared IBD. The basic idea behind this test is that, whenever a marker locus is closely linked to a gene that influences a phenotype, a significant negative regression coefficient should exist between number of alleles shared IBD and the squared phenotypic difference between sibs. This is true because siblings who are phenotypically more alike should share more alleles IBD at the marker locus than sibs who are phenotypically dissimilar. Numerous studies have demonstrated that for quantitative trait analysis the variance component method discussed below is more powerful than methods based on sib pair differences in traits (Amos et al., 1997; Pugh et al., 1997; Williams and Blangero, 1999). Recently, the Haseman-Elston method has been revised (Elston et al., 1998) to use the product of the sibs' trait values rather than the difference. This new method addresses the phenotypic covariance between sibs and is functionally similar to variance component methods except in that it uses regression rather than maximum likelihood. Least squares regression is computationally more efficient than maximum likelihood, and the new Haseman-Elston method will be more rapid than variance component methods. However, the variance component method explicitly deals with the non-independence of the elements of the phenotypic covariance matrix, and since it is likelihood based, it yields a more powerful test than the Haseman-Elston method.

Variance components methods

The variance of a quantitative trait describes the spread of the phenotypic values around the population mean. The basic idea behind the variance component linkage method is to attribute that population variance to a variety of genetic and non-genetic causes. The variance component method seeks to explain the correlations in phenotype among members of a family by partitioning the phenotypic variance into components due to the effect of a specific QTL linked to a genotyped marker, to other QTLs unlinked to the region under consideration, to environmental factors shared among family members such as diet, and to individual-specific environmental sources such as measurement error (Hopper and Mathews, 1982; Amos, 1994). These methods have recently been extended to allow multipoint linkage analysis in pedigrees of arbitrary size and complexity (Almasy and Blangero, 1998) and to oligogenic analysis of multiple QTLs simultaneously (Blangero and Almasy, 1997;

Almasy and Blangero, 1998). For a simple model with n QTLs, the covariance among pedigree members is modeled as:

$$\Omega = \sum_{i=1}^{n} \Pi_i \sigma^2_{qi} + 2\Phi\sigma^2_a + I\sigma^2_e$$

where Π_i is a matrix of IBD sharing among family members at marker i, σ^2_{qi} is the additive genetic variance due to a QTL linked to marker i, Φ is a matrix of kinship values, σ^2_a is the residual additive genetic variance, σ^2_e is the individual-specific environmental variance, and I is an identity matrix. One advantage of this method is that it provides not only an estimate of QTL location, but also an estimate of the magnitude of effect of the QTL, through σ^2_{qi}. The maximum likelihood of a model in which one or more QTL components is estimated is compared to that of a model in which these components are constrained to zero, testing the hypothesis that σ^2_{qi} is significantly greater than zero. Twice the difference in $\log_e$ likelihood between these models is asymptotically distributed as a mixture of chi-square distributions, with the mixing proportions depending on the number of QTL components estimated (Self and Liang, 1987). For a single locus analysis, the difference in the $\log_{10}$ likelihood between the two models is equivalent to the LOD score obtained from classical parametric linkage analyses. These maximum likelihood analyses assume that the phenotype has a multivariate normal distribution, meaning that the plot of trait values follows a Gaussian or bell curve. While the method is robust to violations of normality, such as skewness in the distribution (Amos, 1994), it is sensitive to outliers whose phenotypic values are discontinuous from the rest of the distribution. However, the normality of the distribution and the presence or absence of outliers are easily assessed.

An advantage of the variance component method is that measured covariates, such as age, sex or environmental exposures, are incorporated into the model as modifiers of the phenotypic mean and covariate effects are estimated simultaneously with the components of variance. Incorporation of covariate effects reduces the unexplained variance in the phenotype and increases the proportion of that variance due to the QTLs being sought, thus increasing power to detect linkage.

It is also simple to incorporate more complex genetic models in variance component analyses. These include additional sources of variance such as dominance genetic effects, where the genotypic mean of heterozygotes does not fall halfway between that of the two homozygotes, or shared environmental components common to members of a household such as diet or exposure to pathogens (Almasy and Blangero, 1998). Interactive effects can be accommodated for both gene by gene (epistatic) interactions (Mitchell et al., 1997) and gene by environment interactions (Towne et al., 1997). A common example of gene by environment interaction is genotype by sex interaction, where a gene has different effects in males and females due to the different hormonal environments of the two sexes. The variance component method also can exploit genetic and environmental correlations between phenotypes through multivariate analyses to improve power to detect linkage and improve localization of QTLs (Almasy et al., 1997).

The power of the variance component method depends on the proportion of the phenotypic variance of the trait accounted for by the gene being sought (the QTL-specific heritability), the size of the sample, the distribution of those individuals across families, and whether the sample has been ascertained on individuals with a disease or an extreme quantitative phenotype. As noted earlier, heritability in the broad sense is the proportion of the phenotypic variance that is attributable to all genetic effects. Typically, in the variance components linkage approach, dominance effects are disregarded and additive genetic, or narrow sense, heritabilities are estimated. QTL-specific heritability is the proportion of the phenotypic variance attributable to the additive genetic effects of a specific QTL. Higher QTL-specific heritability and larger sample size yield greater power to detect linkage. Generally, the larger the sampling unit, the more powerful the variance component analysis (Williams et al., 1997). On a per person

basis, nuclear families provide more linkage power than sib pairs and extended pedigrees provide more power than nuclear families. Whereas theoretical power calculations can be made for nuclear families of specified size, quasi-analytical power calculations on extended pedigrees require some empirical simulation (Williams and Blangero, in press).

It is important to remind the reader that the immediately preceding points concerning statistical power apply to the detection of genetic effects on normal variation in *common*, continuously varying, complex traits. Such normal variation is the primary focus of this review, as well as of most anthropological research. However, for *rare* traits — such as the rare autosomal recessive Dubin-Johnson syndrome (hyperbilirubemia II) or methemoglobinemia due to methemoglobin reductase deficiency — methods that combine stringent ascertainment schemes and data collected from smaller sampling units are predictably more successful than those employing random ascertainment in large, extended pedigrees. Approaches more suitable to the analysis of discrete traits with low prevalence are touched upon below in the next section of this review.

DISCRETE TRAITS

Common discrete traits

The variance component linkage method has recently been extended to accommodate analysis of dichotomous traits by use of a threshold model (Duggirala et al., 1997). The threshold model assumes that an individual will exhibit a specific disease state or phenotype if an underlying genetically determined risk or liability exceeds a certain threshold on a normally distributed liability curve. Unaffected individuals are assumed to have underlying liability values below the threshold, and the threshold is set according to the disease or phenotype prevalence so that a proportion of the liability curve equal to the prevalence lies above the threshold. The liability is assumed to have an underlying multivariate normal distribution, but as with quantitative traits the test is robust to non-normality. The correlation in liability between pairs of individuals is estimated using the affection status of unrelated individuals and various classes of relatives. For a simple model, the correlation in liability (ρ) between individuals i and j is given by:

$$\rho_{ij} = \hat{\pi}_{qij} h_q^2 + 2\phi_{ij} h^2 + I_{ij} e^2,$$

where $\hat{\pi}_{qij}$ is the proportion of alleles that individuals i and j share IBD at a QTL (q); h_q^2 is the QTL-specific heritability; h^2 is the residual additive genetic heritability; I_{ij} is the coefficient for the individual-specific environmental component; and e^2 is $1 - h^2 - h_q^2$ (i.e., the remaining variance not due to genetic effects). Measured covariates that may affect liability are incorporated into these analyses using a threshold for affection that varies with the vector of covariates. Thus males and females or smokers and nonsmokers may have different prevalences of disease.

The power of the variance component method for discrete traits is dependent on the sample size and QTL heritability, as for quantitative traits, but also on the relative informativeness of affected and unaffected individuals with regard to the unobserved liability distribution. If the population prevalence of the disease or phenotype is 1%, the risk threshold on the unobserved quantitative liability distribution is placed such that 99% of the curve is below the threshold. Thus unaffected individuals could have a wide range of true liability values and knowing that an individual is unaffected reveals very little about their unobserved liability. On the other hand, knowing that an individual is affected severely restricts their potential liability values. In these cases, when phenotype prevalence is very low, unaffected individuals provide little information and analyses restricted to affected individuals, such as those described in the next section, will be most powerful. However, when the population prevalence is high, unaffected individuals can provide as much information as affected individuals. The power of the discrete trait variance component method is maximized as prevalence approaches 50% and theoretical calculations suggest that it will be more efficient than affected-only methods when the prevalence is over 10% (J.T. Williams and J. Blangero, unpublished data).

Rare discrete traits

The relative pair method is based on detecting deviations in IBD allele sharing from that which is expected based on a relative pair's kinship (Suarez et al., 1978; Risch, 1990). The basic idea is similar to that of the Haseman-Elston method (Haseman and Elston, 1972) described above. That is, relatives who are phenotypically alike (e.g., concordant for phenotype) should share more alleles IBD at a marker locus linked to a QTL than is expected by chance. This can be parameterized in terms of mean IBD allele sharing ($E(\pi)$) or in terms of the expected proportion of relative pairs sharing 0, 1, or 2 alleles IBD (z_0, z_1, and z_2). Sibling pairs are the most commonly studied relative pairs, with expectations of ½ for π and ¼, ½, and ¼ for z_0, z_1, and z_2, respectively. Phenotypically concordant sib pairs (whether affected or unaffected) are expected to share more than half of their alleles IBD, whereas phenotypically discordant sib pairs are predicted to share less than half of their alleles IBD at markers linked to a QTL. Similarly, concordant sib-pairs are expected to share two alleles IBD more than one-quarter of the time, whereas discordant pairs are expected to share two alleles IBD less often. An important parameter for relative pair analyses is the relative risk, λ_R, which is K_R/K, where K_R is the risk of disease (or phenotype) in a relative of type R and K is the population prevalence. The relative risk is a measure of how much more likely a relative of an affected person is to exhibit the phenotype as compared to a randomly chosen individual from the population. Just as heritability can be discussed either in QTL-specific terms or more generally for unspecified QTLs, the relative risk can also be considered in general for the trait, or in relation to specific QTLs. The proportion of alleles an affected relative pair is expected to share IBD at a QTL influencing disease status is a function of the relative risk above population prevalence that is attributable to that QTL.

A number of test statistics have been developed to evaluate the significance of relative pair deviations from IBD expectation. A one-sided test based on normal deviates may be used to assess deviations of z_2 from ¼ in sib-pairs (Day and Simmons, 1976; Suarez et al., 1978) or deviations of π from its expectation in any type of relative pair. Risch (1990) has developed a maximum likelihood-based test that compares the likelihood of the observed IBD distribution given predicted IBD sharing at a linked QTL versus the likelihood of the observed IBD data given expected IBD sharing under a recombination fraction of ½ (i.e., no linkage).

The power of affected relative pair methods depends on the sample size and the population prevalence of the phenotype. Risch (1990) has demonstrated that the power of affected pair methods is also a function of the relative risk, λ_R. In order to calculate the power to detect linkage for a given sample size and λ_R, assumptions must be made about the number of QTLs which exist and the interactions between them.

AFTER THE INITIAL LINKAGE

Once an initial positive linkage result has been obtained, the next step is to narrow the genomic region of interest. A typical QTL linkage analysis using a 10 cM linkage map will often produce a candidate region 20–30 cM wide. On average across the human genome, 1 cM corresponds to about 1 megabase of DNA sequence. Given the density of functional genes per megabase, a region of 20 cM may contain more than 100 functional protein-coding loci. To move from an initial linkage result to identification of the one functional gene — the QTL — that is influencing phenotypic variation, it is necessary to reduce the chromosomal area of interest.

The general practice following initial positive genetic linkage results is to saturate the chromosomal segment with additional microsatellite markers. Given the number of microsatellites known in the human genome, it is straightforward to add markers to the region until the average spacing is less than 3–5 cM. This added linkage information can narrow the area believed to contain the QTL. But it is unlikely that linkage analysis using current methods can reduce the chromosomal region of interest to much less than about 5 cM. These methods are not capable of distinguishing the effects of two adjoining

chromosomal segments that are separated by recombination in only 2 or 3% of meioses.

After using saturation mapping to reduce the area of interest, researchers will generally attempt to exploit linkage disequilibrium as the next step in QTL localization. Any two linked polymorphisms will show some pattern of association within a given population. If there is a random distribution of alleles at two or more loci (or polymorphisms within a locus) across a chromosomal segment, then the polymorphisms are said to be in linkage equilibrium. If, on the other hand, alleles of one polymorphism are found on the same chromosome with specific alleles of a different polymorphism more frequently than is expected by chance, then the polymorphisms are said to be in linkage disequilibrium. This disequilibrium will develop as a result of the specific history of this chromosomal segment in the particular population under study. The history of new mutation, recombination, genetic drift, and selection for specific alleles or combinations of alleles, will determine the pattern of linkage disequilibrium observed in any one DNA segment at any one point in time.

Efforts to identify QTLs exploit linkage disequilibrium by searching for nonrandom associations of genetic markers (presumably neutral polymorphisms) and the true functional polymorphism that influences a phenotype of interest. In association studies, individual polymorphisms, such as single nucleotide polymorphisms (SNPs) or insertion/deletion polymorphisms, are genotyped in a set of individuals. Statistical association between alleles or genotypes and the phenotype under study is tested by looking for mean phenotypic differences among groups of individuals classified by marker genotype. A significant difference among genotypes is a positive association.

This is a valuable approach to identifying the functional gene underlying the QTL effect, because linkage disequilibrium will decay over relatively short distances across the genome. Polymorphisms separated by 150,000–200,000 base pairs are unlikely to show significant linkage disequilibrium. However, the pattern of linkage disequilibrium within human chromosomal regions is poorly understood at this time, and the available data indicate that disequilibrium can vary substantially among a series of polymorphisms within a small region (Clarke et al., 1998; Terwilliger and Weiss, 1998). False positive associations can appear for a variety of reasons, and there is no guarantee that randomly selected genetic markers in the correct functional locus will in fact exhibit linkage disequilibrium with the functional mutation. Unfortunately, no more powerful method is currently available to test the relationship between a specific candidate gene and the inferred phenotypic effect. Microsatellites are not often used in association studies because their high mutation rate will tend to reduce linkage disequilibrium between a functional mutation and microsatellite alleles more rapidly than between a functional mutation and SNPs or other types of nucleotide polymorphism. Overall, association tests are an important tool, but they are subject to high rates of false positives and false negatives.

Both the scientific and the lay press have heralded the upcoming completion of the Human Genome Project and lead us to expect a catalogue of all human genes within the next 5 years. Even though many of their functions and interactions with other genes and environmental factors will not be known initially, the ability to quickly obtain from computer databases a complete list of all the functional genes in any given chromosomal segment will dramatically and permanently change human genetic analyses, especially QTL linkage mapping. Those genes that are located near the peak linkage signal for a given phenotype will become valuable positional candidate loci. Polymorphic markers within these positional candidates can be identified, and in some cases will already be described in public databases. Such polymorphisms will be used to refine a linkage analysis and better localize and identify a detected QTL. Ultimately, the goal of all QTL mapping studies is to identify the true functional mutation or mutations that underlay observed phenotypic variation. Accomplishing this goal remains a formidable task, involving the methods discussed above, as well as molecular experiments designed to determine the exact phenotypic consequences of specific changes in gene structure

or expression. These latter analyses (e.g., studies of the functional consequences of amino acid variation in proteins, gene regulation, differential patterns of gene expression, etc.) are beyond the scope of this review.

OBESITY: AN EXAMPLE PHENOTYPE IN HUMANS

As mentioned in the introduction, few examples of QTL mapping relevant to biological anthropology have yet been published. However, body composition is one trait that has received substantial attention, and is of basic interest to human biology and anthropology. Classical anthropometric studies long ago documented individual and population level differences among humans in adiposity, body mass index (BMI) and other indices of body composition. It is also well established that obesity is a complex trait, influenced by both genetic and environmental factors (Comuzzie and Allison, 1998). It is therefore an example of the type of trait that can be investigated through QTL mapping approaches.

At this point, the results of four genome scans for obesity-related phenotypes have been published (Comuzzie et al., 1997; Norman et al., 1997, 1998; Lee et al., 1998; Hager et al., 1998) and several others are nearing completion. While these published results are from a variety of populations (e.g., Mexican American, Native American, and Western European), some consistent patterns have begun to emerge. At present the strongest evidence for a quantitative trait locus influencing obesity-related phenotypes in humans comes from the San Antonio Family Heart Study (SAFHS) (Comuzzie et al., 1997). This study utilized a sample of 459 Mexican Americans distributed in ten families (this sample represents 5667 relative pairs ranging from parent-offspring to double second cousins) using a variance component linkage approach (Almasy and Blangero, 1998) to genomic screening. Results were initially reported based on a 20 cM map of microsatellite markers, and identified significant linkages on chromosomes 2 (LOD = 4.95) and 8 (LOD = 2.2) for leptin levels and a significant linkage for fat mass (FM) on chromosome 2 (LOD = 2.75) (Comuzzie et al., 1997). Leptin is a circulating protein secreted by adipocytes in proportion to the total amount of body fat present (Freidman and Haalas, 1998) and is a valuable index of adiposity. Following the typing of additional markers to create a 10cM map, the multipoint LOD score for leptin on chromosome 2 increased to 7.46 (Figure 2), the largest LOD score yet published for a QTL in humans (Hixson et al., 1999). Thus far, this is the only QTL linkage result in humans with published replications (Rotimi et al., 1999; Hager et al., 1998). In addition to the leptin linkage on chromosome 8, a significant linkage (LOD = 3.21) was detected in this same region for BMI (Mitchell et al., 1999). An important aspect of the variance component approach to linkage analysis is the fact that it not only allows for the localization of genes, but also provides an estimate of the magnitude of their effect on the phenotype being analyzed (see above). In the case of the linkage results for chromosome 2, this QTL is estimated to account for 47% of the variation in serum leptin levels and 32% of the variation in fat mass in Mexican Americans (Comuzzie et al., 1997). The same locus seems to account for 56% of the variation in leptin levels in African Americans (Rotimi et al., 1999).

In the case of both the chromosome 2 and chromosome 8 linkage results, strong candidate genes for obesity have been found within the chromosomal regions identified. The 95% confidence interval surrounding the chromosome 2 QTL contains the *POMC* locus, which codes for the prohormone pro-opiomelanocortin. *POMC* was identified as a candidate based on its location (Comuzzie et al., 1997), and is further supported by recent studies detailing its potential physiological involvement in appetite regulation (Woods et al., 1998). Recently completed work has now identified polymorphisms in *POMC* that can be used in formal association analyses. With a haplotype generated using two common polymorphisms in *POMC* (one located in exon 3 and the other in the 5′ untranslated region) we have detected significant association ($p = 0.001$) between molecular variation in the *POMC* locus and variation in serum levels of leptin among Mexican Americans (Hixson et al., 1999). Since neither of

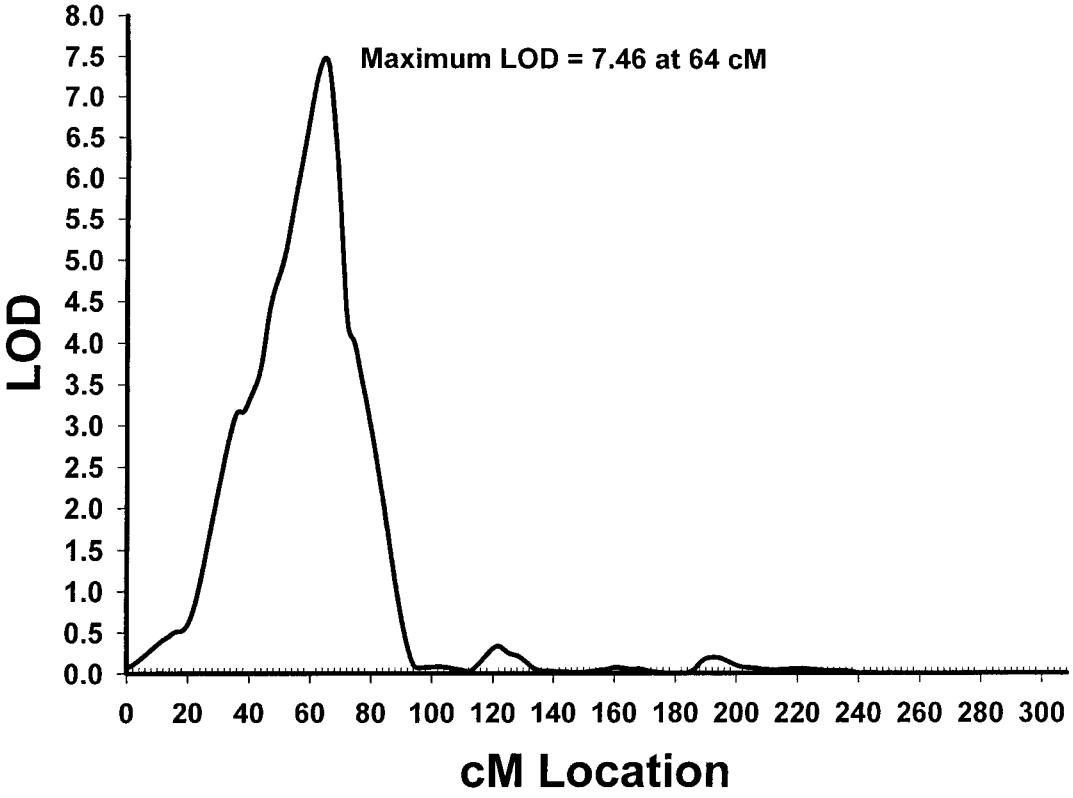

Fig. 2. Estimated LOD score functions obtained from multipoint quantitative trait linkage analysis of serum leptin levels in Mexican Americans. The horizontal axis reflects position along chromosome 2.

these two polymorphisms appears to be functional, work is currently focusing on identifying additional polymorphisms in the promoter region of *POMC*.

The 95% confidence interval surrounding the chromosome 8 linkage contains the structural gene for the β-3-adrenergic receptor (*ADRB3*), a strong candidate gene previously identified based on its known physiological activity. Given these linkage results, *ADRB3* is supported as a potential contributor to observed variation on the basis of chromosomal position as well. While previous association studies testing the relationship between the Trp64Arg polymorphism of *ADRB3* and obesity-related phenotypes yielded equivocal results (Allison et al., 1998), the argument for *ADRB3* as a human obesity gene has been strengthened by follow-up analyses in this same sample of Mexican Americans discussed above with regard to *POMC* (Mitchell et al., 1998). By first accounting for the effects of the QTL identified on chromosome 2, Mitchell and colleagues (1998) were then able to detect association between variation in *ADRB3* and several obesity-related phenotypes (i.e., BMI, FM, waist circumference). This finding is particularly interesting since it exemplifies one of the ways in which positive results from genomic screens in one region of the genome can be used to improve the likelihood of discerning the effects of other candidate genes in other genomic regions.

BEHAVIORAL GENETICS

Human behavioral phenotypes, which have engendered much controversy and fueled the nature versus nurture polemic, are also yielding to new analytical strategies and methods developed for the genetics of complex traits. In studies of *common* behavioral traits or common quantitative traits associated with behaviors and/or behavioral

disorders, it is now clear that analyses of data from extended pedigrees provide a bonus of statistical power over analyses of smaller sampling units, such as sibships. These sorts of studies also provide the researcher with the ability to differentiate correlations among siblings due to rearing environment from those due to QTLs by incorporating many distantly related relatives who share genes, but not environment. A positive linkage finding cannot be caused by shared environmental influences as the observed allele sharing at a genotyped marker provides precise predictions concerning the expected correlations among relatives in excess of those attributable to kinship alone, and these predictions do not mimic shared environment. On the basis of genotype alone, we may predict that an individual will be more like a distant relative, whom he or she has never met, than like a sibling. A third cousin may share an allele IBD at the QTL while a sibling may share no alleles IBD at the same QTL. Over a large pedigree, this can generate the statistical support for a linkage relationship between a given behavioral phenotype and microsatellite markers in a specific chromosomal region. The same observation can be made in sibship-based analyses where QTL genotype predicts which sibs will be more alike and which less alike.

The Tridimensional Personality Questionnaire (TPQ) provides quantitative measures of three aspects of personality: harm avoidance, reward dependence, and novelty seeking. Using TPQ data from extended pedigrees ascertained through an alcoholic proband as part of the Collaborative Study on the Genetics of Alcoholism (COGA) (Begleiter et al., 1995), a number of variance components linkage analyses have been conducted. Cloninger et al. (1998) localized a QTL for harm-avoidance, thought to be related to anxiety-proneness, to chromosome 8 with a LOD score of 3.2. Czerwinski et al. (1999) employed TPQ measures to better inform a linkage analysis of alcoholism in 660 individuals from these same pedigrees. Their analysis included the continuous TPQ subscale measure for novelty-seeking behavior in a bivariate mixed (i.e., continuous and dichotomous) linkage analysis with the discrete trait alcoholism. The results demonstrated that this bivariate approach substantially improved the evidence for a QTL on chromosome 4 influencing risk of alcoholism.

Other complex human behavioral traits that are currently the subject of linkage screens include alcohol dependence (Reich et al., 1998), reading disability (Cardon et al., 1994; Grigorenko et al., 1997), and sexual preference (Hamer et al., 1993; Wickelgren et al., 1999). Many of these studies remain controversial and are still awaiting follow-up in the form of independent replications or studies of linkage disequilibrium within candidate regions.

QTL MAPPING IN NONHUMAN PRIMATES

More than 10,000 polymorphic genetic markers have been identified and mapped in the human genome. These loci can be readily genotyped in large numbers of people for use in QTL linkage analyses. As discussed above, the statistical power of any linkage analysis depends in part on the density of highly informative markers in the linkage map. QTL mapping methods can be used in other species as well, and high density genetic linkage maps are available for mice (http://www.informatics.jax.org), rats (Jacob et al., 1995, http://ratmap.gen.gu.se), cattle (Berendse et al., 1994, http://bos.cvm.tamu. edu/bovgbase.html), sheep (deGotari et al., 1998), dogs (Mellersh et al., 1997) and other agricultural animals (Andersson et al., 1996).

The mapping of nonhuman primate genomes has not progressed as rapidly. One species that has been investigated in detail is the baboon (*Papio hamadryas*). Using over 300 microsatellite loci first cloned from the human genome, a linkage map with average spacing between loci of less than 8 cM has been developed (Rogers et al., 1995; Perelygin et al., 1996; Morin et al., 1998; Rogers et al., submitted). This baboon linkage map provides opportunities to use variance component linkage methods to localize genes that influence any phenotypic trait that varies within this species. Research projects are underway to map genes related to bone density (Mahaney et al., 1997), serum cholesterol levels (Mahaney et al., 1998), and other disease-related phenotypes. Future analyses are envisioned that

will address a wider range of phenotypes, including normal variation in traits not related to specific diseases.

The basic strategy employed to construct this baboon linkage map was to use information and resources generated by the Human Genome Project. In a collaboration involving Phil Morin and others from Axys Pharmaceuticals, Inc. (La Jolla, CA), as well as researchers from the Southwest Foundation, human microsatellite loci were screened to identify highly informative microsatellite polymorphisms in the pedigreed colony of baboons maintained at the Foundation. These polymorphic loci were then analyzed in about 700 baboons from several multi-generation families. Because the loci analyzed in the baboons have already been mapped in the human genome, it is straightforward to compare homologous chromosomal segments in the two species. The same approach could also be used in other nonhuman primates. Human microsatellites that are also polymorphic in rhesus monkeys have been identified by Morin et al. (1997) and others. Similar markers for chimpanzees (Deka et al., 1994), pigtailed macaques (Nair et al., submitted for publication), vervet monkeys (Newman et al., 1998) and squirrel monkeys (Witte and Rogers, 1999) have also been described. The greater genetic distance separating New World monkeys from humans, as compared to the divergence of Old World monkeys from humans, makes this approach less effective for the platyrrhines than for catarrhines (Witte and Rogers, 1999). Fewer human PCR primers function well in the more distantly related taxa. For New World monkeys and other primates more divergent from humans, it will probably be necessary to clone novel microsatellites from the species under study in order to generate large numbers of polymorphic markers.

If genetic linkage maps were developed for a number of primate species, these animals could become tremendously valuable to QTL mapping programs in biological anthropology. Few captive colonies of primates have large, multi-generation pedigrees immediately available for study. But molecular genetic markers, especially microsatellites, can be used to reconstruct paternity and confirm matrilineal relationships (Morin et al., 1994; Altmann et al., 1996). Once appropriate pedigrees are available, and genetic linkage maps developed, then any phenotype of interest could be investigated in those nonhuman pedigrees. Environmental factors such as diet and exposure to pathogens are carefully controlled, and consequently the environmental variance of complex traits may in some cases be lower in the pedigreed monkey colonies than in analogous human pedigrees. This can increase the statistical power of QTL mapping projects. Furthermore, various types of longitudinal studies, such as annual X-rays for measurement of bone growth and change, or repeated collection of cerebrospinal fluid for neurochemical analyses, can be done in primates but are not feasible in large human families. Some morphological traits of interest to biological anthropology may be variable in a given primate species but not in modern humans (e.g., third premolar crown morphology in Cebids; metrics of features associated with molar bilophodonty and external tail length in Cercopithecines; sagittal ridge morphometrics in gorillas, etc.). This makes the monkeys or apes the only option for mapping QLTs for these traits.

CROWN-RUMP LENGTH: A LINKAGE ANALYSIS EXAMPLE IN PEDIGREED BABOONS

We have conducted a statistical genetic analysis of crown-rump length, a metric trait often assessed in population studies of humans and nonhuman primates. This will serve as an illustration of the application QTL mapping to traits of interest to biological anthropologists. In the course of investigating the genetics of bone mineral density in baboons, we measured crown-rump length in 713 olive and yellow baboons maintained at the Southwest Foundation for Biomedical Research. These animals, 482 females and 231 males between the ages of 5.5 and 30 years, were assigned to 12 extended pedigrees for the purposes of this analysis. While all 713 animals can be organized into a single, larger extended pedigree, this sample of animals was "broken" into 12 smaller, extended pedigrees to decrease the numerical and computational complexity to a more

manageable level. It was reasonable to expect that both genes and environmental factors influence variation in crown-rump length in baboons. It also was reasonable to expect that the heritable component of variation in a trait such as crown-rump length may represent variation in the sequence and/or expression of several genes. Like many continuously varying quantitative traits, the crown-rump length measure essentially sums the metrics of multiple structures, including vertebrae, intervertebral disks, and cranial skeletal elements that have resulted from an ontogenetic progression of determination, differentiation, growth, development and aging. Each of these stages is influenced to some extent by genes. The number of loci contributing to variation in a trait is inversely related to their mean effect size, and our power to detect linkage to a QTL is a function of the effect size of that QTL. Consequently, for traits like crown-rump length — for which we have a reasonable a priori expectation of the involvement of many loci — the probability that one or more QTLs can be detected and localized is enhanced when the additive genetic component to the variance in the trait is larger rather than smaller.

Initial maximum likelihood based variance components analysis found that sex, sex-specific age terms, and body weight accounted for more than 57% of the total phenotypic variance in crown-rump length in these baboons ($p \ll 0.000001$). Just over 10% of the phenotypic variance is due to variation in body weight alone ($p \ll 0.000001$). Of the residual 43% of the phenotypic variance in crown-rump length that is not explained by these covariates, 56% was attributable to the additive effects of genes ($h^2 = 0.56$, $p < 0.000001$) and 44% to the effects of unmeasured environmental factors. We reparameterized this quantitative genetic model by allowing for a QTL to contribute to the variance in crown-rump length, and we conducted a variance components linkage screen in which we estimated the proportion of the variance in crown-rump length attributable to each of more than 300 microsatellite marker loci distributed across the 20 baboon autosomes. While this screen yielded ten LOD scores >1.0 on six chromosomes, only a marker that maps to human chromosome 7, D7S524, returned a LOD of sufficient magnitude, LOD = 2.49, to be considered *suggestive* evidence (Kruglyak and Lander, 1995a,b) for a QTL for crown-rump length in baboons.

Thorough and careful genetic characterization of a complex trait like crown-rump length can have important implications for attempts to detect and localize QTLs. The reader will remember that body weight exhibited a significant effect on variation in our example trait, accounting for about 10% of the phenotypic variance. When the value for the body weight parameter was fixed at zero (the null model in the likelihood ratio test of the significance of body weight), the estimated residual heritability increased by nearly 5%, from $h^2 = 0.56$ to $h^2 = 0.59$. Because the addition of body weight to the model decreases rather than increases the estimated heritability, this suggests that (1) these two traits may be genetically correlated and (2) the addition of this covariate to the linkage model may compromise our attempts to detect QTLs for crown-rump length.

A bivariate quantitative genetic analysis supports the hypothesized genetic correlation between the two traits (M.C. Mahaney, unpublished results) and a second two-point, whole genome linkage screen using a variance component model that does not include body weight as a covariate illustrates the importance of maximizing the genetic signal to noise ratio. This screen returned fourteen LOD scores > 1.0 on nine chromosomes. More important, it also resulted in increased LOD scores on three chromosomes, 5, 7, and 12, and identified two additional regions on chromosomes 5 and 12 with suggestive evidence of linkage.

Multipoint linkage analyses were employed to simultaneously exploit information on allele sharing between relatives at multiple loci along the three chromosomes where we had observed suggestive evidence of linkage in our initial two-point screen. In our experience, the proper application of multipoint linkage analysis can substantially improve and better localize a valid linkage signal while eliminating a false positive one. Based on the results of our multi-

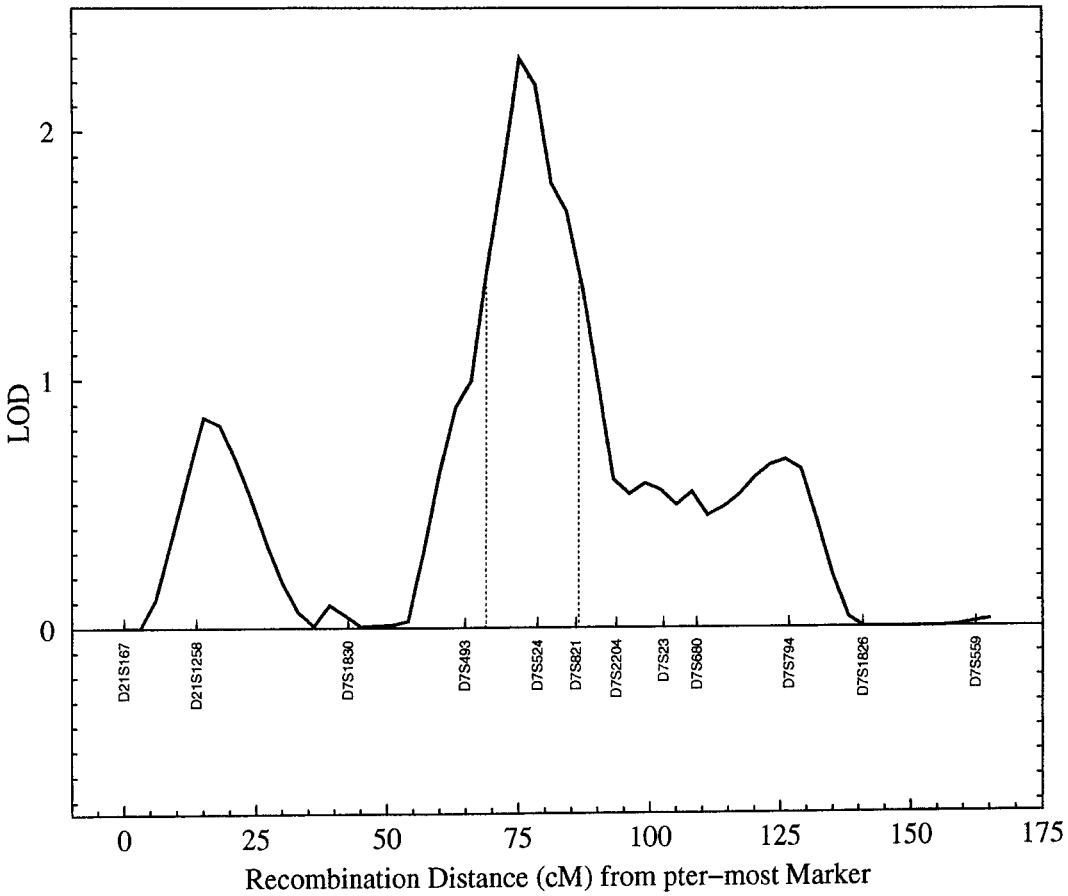

Fig. 3. Multipoint linkage analysis in pedigreed baboons: Crown-rump length and highly polymorphic microsatellite marker loci that map to human chromosome 7 (baboon chromosome 3, fusion of human 21 and 7). Vertically oriented alpha-numeric labels indicate marker loci and their positions in the baboon genome map. Vertical dotted lines indicate the pter (left) and qter (right) boundaries of the 95% confidence interval for the suggested QTL.

point linkage screens the evidence for a possible QTL for baboon crown-rump length on chromosome 7 remains only suggestive (maximum multipoint LOD = 2.29), while that for linkage to chromosomes 5 and 12 is diminished substantially (maximum multipoint LOD < 1.0). If, for the purposes of discussion, we accept the hypothesis that chromosome 7 harbors a QTL for crown-rump length, then the plot of the multipoint LOD scores in Figure 3 suggests that this QTL is located approximately 75 cM (with a 95% confidence interval $\pm$ 9 cM) in the direction of the chromosome's long arm, or q-arm, terminus (qter) from the marker in our map that is nearest the short arm, or p-arm, terminus (pter). The preliminary estimate of the proportion of the residual phenotypic variance in crown-rump length attributable to the QTL on chromosome 7 is approximately 0.33, or nearly 56% of the estimated total additive genetic variance and about 16% of the total phenotypic variance for this trait in these animals.

These preliminary analyses in pedigreed baboons have not confirmed that a QTL for crown-rump length exists on the baboon homologue of human chromosome 7. Rather, they have provided a more informed hypothesis regarding where one should focus subsequent efforts to find one of the genes contributing to variation in this complex trait. Such

efforts might be directed toward the better characterization of the trait itself, increasing the pedigree sample size, adding more marker loci to reduce the inter-locus gaps in the region of interest on chromosome 7, and/or reconfirming the validity of the chromosomal locations of all the marker loci in the baboon genomic map. All of these factors can affect the statistical power of current analyses to localize QTLs, the necessary first step to gene identification. If the chromosome 7 QTL is confirmed, we would conclude from this preliminary analysis that approximately 44% of the additive genetic variance in crown-rump length in our baboon population still remains to be accounted for. If this percentage is attributable to one or two additional QTLs, several currently available strategies (Almasy and Blangero, 1998) can be employed successfully to localize them. If, however, three or more additional QTLs are involved, the variance attributable to one or more of them may be insufficient to allow successful localization through this approach.

QTL MAPPING AND FUTURE ANTHROPOLOGICAL RESEARCH

Much of biological anthropology is directed toward a single, overarching objective: to understand or explain human variation in space and time. In our attempts to achieve this understanding, biological anthropologists typically address two fundamental types of research questions: (a) questions of adaptation and (b) questions of history. In studies of adaptation, investigators describe and investigate characteristics of humans and/or nonhuman primates, either living or extinct. The goals of these analyses are generally to identify characteristic biological features (e.g., anatomy, physiology, growth, development, behavior, etc.) both shared and unique among primate species, and to determine how natural selection may have shaped the patterns of inter- and intra-specific variation exhibited by those features. Functional anatomy, analyses of diet and ecological relationships, and interpretations of the ecological or demographic factors determining or constraining social organization are all examples of attempts to address questions of adaptation.

In studies of history, investigators attempt to reconstruct the temporal series of events that have occurred during human or nonhuman primate evolution. A number of different approaches are used, including but not limited to studies of the fossil record, molecular phylogenetics, and human population dynamics. Obviously, these two lines of research (history and adaptation) are often pursued in parallel. Sometimes an investigator will examine them simultaneously in a single research project.

There can be no doubt that analyses of genetic data have become a major component of biological anthropology. Studies of molecular phylogeny have contributed much to our understanding of evolutionary relationships among living taxa. Studies of human population genetics have had a significant impact on our understanding of modern human origins and dispersal. While anthropological genetics has contributed much to the reconstruction of evolutionary history, it has contributed little to the study of primate or human adaptation. We know of no anatomical, physiological, or behavioral adaptation of significance to anthropology that can, at this time, be described in molecular genetic terms. That is, specific genes that contribute to variation in the embryological development of functionally adaptive traits among primates are virtually unknown. Traits reflecting variation in cranial and dental morphology; musculoskeletal anatomy, physiology, and its relationship to patterns of locomotion; or the psychobiology of primate social behavior have all been investigated in great detail and theories concerning their phylogenetic and adaptive significance abound. But the available information includes nothing about the particular genetic differences that contribute to interpopulational and/or interspecific variation in these characteristics. This is a particularly troubling gap in knowledge, given that all adaptive and phylogenetic scenarios are built upon assumptions concerning the mechanisms of inheritance and patterns of genetic variation underlying the traits of interest.

The genetic approaches outlined above provide a means to expand the scope and contributions of anthropological genetics. It is now practicable, and indeed timely, for biological anthropologists to formally ad-

dress questions concerning the genetic bases of normal variation in growth, development, aging, and function in humans and nonhuman primates. The results of such investigations will likely provide, for the first time, molecular genetic bases for the generation of testable hypotheses concerning the evolutionary processes that may have shaped observable patterns of biological variation among primates in general, and our species in particular.

Until recently, there was no practical method available for the identification of individual genes that influence variation in complex traits of interest to biological anthropologists. But advances in both molecular and statistical genetics, made over the past two decades, have provided a set of strategies and procedures that make such studies feasible today. Most of the work done in this area has been motivated by an effort to understand and cope with the genetic basis of human diseases (see Table 2 for useful resources). Even so, as we have attempted to show above, it has direct application to the dissection of the sources of normal biological variation within species and the identification of genes whose expression may contribute to inter-specific variation as well.

While the methodologies for QTL localization reviewed in this paper do have direct application to biological anthropology, it must be noted that they also are associated with relatively high costs that may seem, at first glance, prohibitive to many in this discipline. These costs include, but are not limited to, initial capital outlay for laboratories and equipment; salaries for trained personnel; funding for sample collection, processing, and storage; and funding for both equipment and personnel responsible for the organization, management, analysis, and interpretation of data in volumes that are large, even by the standards of most anthropologists. For a whole genomic screen of a moderate size sample of families or extended pedigrees, these costs can easily be in the millions of dollars, even without accounting for the costs of recruitment of families of participants and the collection of phenotype data.

There are two obvious strategies for biological anthropologists who currently are not funded to conduct such studies, but who would wish to be. The first involves recasting their population-based, anthropological research in biomedical terms and applying for support from agencies with greater re-

TABLE 2. Useful internet resources

Center for Medical Genetics (Marshfield, WI USA)	http://www.marshmed.org/genetics/
Cooperative Human Linkage Center	http://www.chlc.org/
GENATLAS QUERY	http://bisance.citi2.fr/GENATLAS/
Genetics-related resources	http://linkage.rockefeller.edu/outside/list.html
Hopkins Bio-Informatics Home Page	http://www.bis.med.jhmi.edu/
OMIM Home Page—Online Mendelian Inheritan.	http://www3.ncbi.nlm.nih.gov/Omim/
The Genome Database	http://gdbwww.gdb.org/
The Human Genetic Analysis Resource	http://darwin.cwru.edu/
Genetic linkage analysis	http://linkage.rockefeller.edu/
Map Manager Family	http://mcbio.med.buffalo.edu/mapmgr.html
Welcome to GÉNÉTHON	http://www.genethon.fr/genethon en.html
The Human Transcript Map	http://www.ncbi.nlm.nih.gov/SCIENCE96/
GeneMap'98	http://www.ncbi.nlm.nih.gov/genemap99/
Jackson Laboratory Mouse Genomics	http://www.informatics.jax.org
FlyBase @ flybase.bio.indiana.edu	http://flybase.bio.indiana.edu:82/
Galton Laboratories Genetic Maps	http://www.gene.ucl.ac.uk/chr9/genetic.shtml
USC Genetic Epi Genomic Data	http://norp5424b.hsc.usc.edu/genedata.html
Stanford University—Morrison Institute—The Human Genome Diversity Project	http://www.stanford.edu/group/morrinst/HGDP.html
National Human Genome Research Institute (NHGRI)	http://www.nhgri.nih.gov/
Genome Research	http://www.er.doe.gov/production/ober/hug top.html
An STS-based map of the human genome	http://carbon.wi.mit.edu:8000/cgi-bin/contig/physmap
LLNL Human Genome Center	http://www-bio.llnl.gov/bbrp/genome/genome.html
Genetic maps of the rat genome	http://waldo.wi.mit.edu/rat/public/mar97/
Human-mouse homology relationships	http://www.ncbi.nlm.nih.gov/Homology/
The Zebrafish Server	http://zfish.uoregon.edu/
The genetic location database	http://cedar.genetics.soton.ac.uk/public_html/
Biological journals and abbreviations	http://arachne.prl.msu.edu/journams/
National Center for Biotechnology Information	http://www.ncbi.nlm.nih.gov/
Human obesity gene maps	http://www.obesite.chaire.ulaval.ca/genes.html

sources than those to which biological anthropologists traditionally appeal. Such agencies have histories of funding projects that are larger in scale and cost than the average study in biological anthropology. Such recasting should not be difficult. Much of what biological anthropologists study has clear implications for the health and welfare of our species. Indeed, such implications are often required considerations in the penultimate sections or paragraphs of study proposals, manuscripts, or podium presentations addressing traditional problems in biological anthropology. What is required is that these biomedical considerations become the central foci of well-considered research proposals that are informed and strengthened by their evolutionary and anthropological components. When completed, these studies will provide the biological anthropologist with the data to address and answer questions of direct interest to their own discipline as well.

The second strategy is to identify funded, ongoing biomedical research projects that have conducted or are in the process of conducting whole-genome searches in populations of interest to biological anthropologists. A number of such studies are currently collecting genotypic and phenotypic data in human populations that are distinguished by culture, socioeconomic status, ethnicity, geography and/or ecology, or some combination thereof. Some examples of large-scale studies of cardiovascular disease and its risk factors (e.g., diabetes, obesity, lipid and lipoprotein metabolism, etc.) are the Phoenix Epidemiology and Clinical Research Branch of the National Institute for Diabetes and Digestive Disorders study of Pima Indian families (Pratley et al., 1998); the Quebec Family Study of data from several hundred French Canadian nuclear families (Rankinen et al., 1999); the Strong Heart Study (Howard et al., 1999) with data from extended pedigrees and nuclear families of Native Americans from Arizona, Oklahoma, and North and South Dakota; and the San Antonio Family Heart Study (Rainwater et al., 1999) of approximately 1500 Mexican Americans in 42 extended pedigrees. Other examples include genome screens focusing on measures of growth, development, maturation, and body composition data from families participating in the Fels longitudinal growth study (Roche, 1992), susceptibility to helminthic infection in a large, extended pedigree of Jirels from Eastern Nepal (Williams-Blangero et al 1998), susceptibility to the Chagas disease parasite, *Trypanosoma cruzi*, in rural Brazilian families (Williams-Blangero et al., 1997), and obesity and diabetes in American Samoa (R. Deka, personal communication).

Biological anthropologists and primatologists interested in localizing and identifying QTLs in nonhuman primate species can look to fewer ongoing genomic screens at the present time. This is due in large part to the fact that a reliable map of a species' genome is a necessary prerequisite to any QTL mapping efforts. When such a map exists, genes that influence complex traits can be located using data from nonhuman primates as we have tried to demonstrate with our example of crown-rump length in baboons. In addition to the ongoing projects to localize and identify QTLs for age-related changes, measures of formation, resorption and pathology in bone, the genome map developed for the pedigreed baboon colony at the Southwest Foundation for Biomedical Research is also being explored for QTLs influencing variation in measures of lipid and lipoprotein metabolism (Mahaney et al., 1998), hemostasis and hematology (Mahaney, unpublished data), and dental size and morphology (Hlusko, unpublished data). A vervet monkey project (Newman et al., 1998) has as one of its long-term goals the localization of QTLs influencing variation in behavior and neurochemistry. While only some of these projects have begun to yield linkage results, each of the phenotypes mentioned above, as well as innumerable others, is amenable to genetic linkage mapping. Studies explicitly designed to map genes for complex traits such as these can contribute substantially to the body of knowledge that is biological anthropology.

ACKNOWLEDGMENTS

Support for this review and for some of the results reported therein was provided in part by the National Institutes of Health (P51 RR13986, P01 HL28972 , P01 HL45522, R01 RR08781, R01 HL54141, R01 MH59490, F32 GM18897), and by Axys Pharmaceuti-

cals, Inc. (formerly Sequana Therapeutics, Inc.). The authors thank Phillip Morin (Axys) and Deborah Newman (SFBR) for contributions to the baboon gene mapping analyses.

LITERATURE CITED

Allison DB, Heo M, Faith MS, Pietrobelli A. 1998. Meta-analysis of the association of the Trp64Arg polymorphism in the beta$_3$ adrenergic receptor with body mass index. Int. J. Obes. 22:559–566.

Almasy L, Dyer, TD, Blangero J. 1997. Bivariate quantitative trait linkage analysis: Pleiotropy versus coincident linkage. Genet. Epidemiol. 14:953–958.

Almasy L, Blangero J. 1998. Multipoint quantitative trait linkage analysis in general pedigrees. Am. J. Hum. Genet. 62:1198–1211.

Altmann J, Alberts SC, Haines SA, Dubach J, Muruthi P, Coote T, Geffen E, Cheesman DJ, Mututua RS, Saiyalel SN, Wayne RK, Lacy RC, Bruford MW. 1996. Behavior predicts genes structure in a wild primate group. Proc. Natl. Acad. Sci. USA 93:5797–5801.

Amos CI. 1994. Robust variance-components approach for assessing genetic linkage in pedigrees. Am. J. Hum. Genet. 54:535–543.

Amos CI, Dawson DV, Elston RC. 1990. The probabilistic determination of identity-by-descent sharing for pairs of relatives from pedigrees. Am. J. Hum. Genet. 47:842–543.

Amos CI, Krushkal K, Thiel TJ, Young A, Zhu DK, Boerwinkle E, de Andrade M. 1997. Comparison of model-free linkage mapping strategies for the study of a complex trait. Genet. Epidemiol. 14:743–748.

Anderson L, Archibald A, Ashburner M, Audun S, Barendse W, Bitgood J, Bottema C, Broad T, Brown S, Burt D, Charlier C, Copeland N, Davis S, Davisson M, Edwards J, Eggen A, Elgar G, Eppig JT, Franklin I, Grewe P, Gill T, Graves JAM, Hawken R, Hetzel J, Hilyard A, Jacob H, Jaswinska L, Jenkins N, Kunz H, Levan G, Lie O, Lyons L, Maccarone P, Mellersh C, Moore S, Moran C, Morizot D, Neff M, Nicholas F, O'Brien S, Parsons Y, Peters J, Postlethwait J, Raymond M, Rothschild M, Schook L, Sugimoto Y, Szpirer C, Tate M, Taylor J, VandeBerg J, Wakefield M, Wienberg J, and Womack J. 1996. Comparative genome organization of vertebrates. Mamm. Genome 7:717–734.

Bailey JN, Almasy L. 1995. A brute force dichotomization approach to quantitative trait linkage analysis. Genet. Epidemiol. 12:719–722.

Barendse W, Armitage SM, Kossarek LM, Shalom A, Kirkpatrick BW, Ryan AM, Clayton D, Li L, Neibergs HL, Zhang N, et al. 1994. A genetic linkage map of the bovine genome. Nature Genet. 6:227–235.

Beall CM, Brittenham GM, Strohl KP, Blangero J, Williams-Blangero S, Goldstein MC, Decker MJ, Vargas E, Villena M, Soria R, Alarcon AM, Gonzales C. 1998. Am. J. Phys. Anthropol. 106:385–400.

Begleiter H, Porjesz B, Reich T, Hesselbrock V, Li T-K, Shuckit MA, Edenburg HJ, Rice J. 1995. The collaborative study on the genetics of alcoholism. Alcohol Health Res. World 19:228–236.

Blangero J, Almasy L. 1997. Multipoint oligogenic linkage analysis of quantitative traits. Genet. Epidemiol. 14:959–964.

Bruford MW, Wayne RK. 1993. Microsatellites and their application to population genetic studies. Curr. Opin. Genet. Develop. 3:939–943.

Buetow KH, Edmonsom MN, Cassidy AB. 1999. Reliable identification of large numbers of candidate SNPs from public EST data. Nature Genet. 21:323–325.

Cardon LR, Smith SD, Fulker DW, Kimberling WJ, Pennington BF, DeFries JC. 1994. Quantitative trait locus for reading disability on chromosome 6. Science 266:276–279.

Cheverud, JM, Routman E, Duarte FAM, van Swinderen B, Cothran K, Perel C. 1996. Quantitative trait loci for murine growth. Genetics 142:1305–1319.

Clark AG, Weiss KM, Nickerson DA, Taylor SL, Buchanan A, Stengard J, Salomaa V, Vartiainen E, Perola M, Boerwinkle E, Sing CF. 1998. Haplotype structure and population genetic inferences from nucleotide-sequence variation in human lipoprotein lipase. Am. J. Hum. Genet. 63:595–612.

Cloninger CR, Van Eerdewegh P, Goate A, Edenberg HJ, Blangero J, Hesselbrock V, Reich T, Nurnberger J Jr, Schuckit M, Porjesz B, Crowe R, Rice JP, Foroud T, Przybeck TR, Almasy L, Bucholz K, Wu W, Shears S, Carr K, Crose C, Willig C, Zhao J, Tischfield JA, Li T-K, Conneally PM, Begleiter H. 1998. Anxiety proneness linked to epistatic loci in genome scan of human personality traits. Am. J. Med. Genet. 81:313–317.

Comparative Gene Mapping. Special issue of ILAR Jounral Volume 39, pages 47–256, 1998.

Comuzzie AG, Allison DB. 1998. The search for human obesity genes. Science 280:1374–7

Comuzzie AG, Hixson JE, Almasy L, Mitchell B, Mahaney MC, Dyer TD, Stern MP, MacCluer JW, Blangero J. 1997. A major quantitative trait locus determining serum leptin levels and fat mass is located on chromosome 2. Nature Genet. 15:273–275.

Comuzzie AG, Williams JT. 1999. Correcting for ascertainment bias in the COGA data set. Genet. Epidemiol. (in press).

Czerwinski SA, Mahaney MC, Williams JT, Almasy L, Blangero J. 1999. Genetic analysis of personality traits and alcoholism using a mixed discrete continuous trait variance component model. Genet Epidemiol. (in press).

Curtis D, Sham PC. 1994. Using risk calculation to implement an extended relative pair analysis. Ann. Hum. Genet. 58:151–162.

Davis S, Schroeder M, Goldin LR, Weeks DE. 1996. Nonparametric simulation-based statistics for detecting linkage in general pedigrees. Am. J. Hum. Genet. 58:867–880.

Day NE, Simons MJ. 1976. Disease susceptibility genes — their identification by multiple case family studies. Tissue Antigens 8:109–119.

de Gortari MJ, Freking BA, Cuthbertson RP, Kappes SM, Keele JW, Stone RT, Leymaster KA, Dodds KG, Crawford AM, Beattie CW. 1998. A second-generation linkage map of the sheep genome. Mamm. Genome 9:204–209.

Deka R, Shriver MD, Yu LM, Jin L, Aston CE, Chakraborty R, Ferrell RE. 1994. Conservation of human chromosome 13 polymorphic microsatellite (CA)n repeats in chimpanzees. Genomics 22:226–230.

Duggirala R, Williams JT, Williams-Blangero S, Blangero J. 1997. A variance component approach to dichotomous trait linkage analysis using a threshold model. Genet. Epidemiol. 14:987–992.

Elston RC, Buxbaum S, Jacobs KB, Olson JM. 1998. Haseman and Elston revisited. Genet. Epidemiol. 15:546.

Falconer DS. 1981. Introduction to quantitative genetics, 2nd ed. Essex: Longman House.

Freidman J, Halaas J. 1998. Leptin and the regulation of body weight in mammals. Nature (London) 395:763–770.

Geldermann H. 1975. Investigations on inheritance of quantitative characters in animals by gene markers. I. Meth. Theoret. Appl. Genet. 46:319–330.

Génin E, Clerget-Darpoux F. 1996. Consanguinity and the sib-pair method: An approach using identity by descent between and within individuals. Am. J. Hum. Genet. 59:1149–1162.

Greenberg DA. 1990. Linkage analysis assuming a single-locus mode of inheritance for traits determined by two loci: Inferring mode of inheritance and estimating penetrances. Genet. Epidemiol. 7:467–479.

Greenberg DA, Hodge SE. 1989. Linkage analysis under "random" and "genetic" reduced penetrance. Genet. Epidemiol. 6:259–264.

Grigorenko EL, Wood FB, Meyer MS, Hart LA, Speed WC, Shuster A, Pauls DL. 1997. Susceptibility loci for distinct components of developmental dyslexia on chromosomes 6 and 15. Am J Hum Genet 60:27–39.

Guo SW, Thompson EA. 1992. A Monte Carlo method for combined segregation and linkage analysis. Am. J. Hum. Genet. 51:1111–1126.

Hager J, Dina C, Francke S, Dubois S, Houari M, Vatin V, Vaillant E, Lorentz N, Basdevant A, Clement K, Guy-Grand B, Froguel P. 1998. A genome-wide scan for human obesity genes shows evidence for a major susceptibility locus on chromosome 10. Nature Genet. 20:304–308.

Hamer DH, Hu S, Magnuson VL, Hu N, Pattatucci AM. 1993. A linkage between DNA markers on the X chromosome and male sexual orientation. Science 261:321–327.

Haseman JK, Elston RC. 1972. The investigation of linkage between a quantitative trait and a marker locus. Behav. Genet. 2:3–19.

Heath SC. 1997. Markov chain Monte Carlo segregation and linkage analysis for oligogenic models. Am. J. Hum. Genet. 61:748–760.

Hixson JE, Almasy L, Cole S, Birnbaum S, Mitchell BD, Mahaney MC, Stern MP, MacCluer JW, Blangero J, Comuzzie AG. 1999. Normal variation in leptin levels is associated with polymorphisms in the pro-opiomelanocortin gene, *POMC*. J. Clin. Endocrinol. Metab. 84:3187–3191.

Howard BV, Lee ET, Cowan LD, Devereux RB, Galloway JM, Go OT, Howard WJ, Rhoades ER, Robbins DC, Slevers ML, Welby TK. 1999. Rising tide of cardiovascular disease in American Indians. The Strong Heart Study. Circulation 99:2389–2395.

Hopper JL, Mathews JD. 1982. Extensions to multivariate normal models for pedigree analysis. Ann. Hum. Genet. 46:373–383.

Krude H, Biebermann H, Luck W, Horn R, Brabant G, Gruters A. 1998. Severe early-onset obesity, adrenal insufficiency and red hair pigmentation caused by POMC mutations in humans. Nature Genet. 19:155–157.

Lucek PR, Ott J. 1997. Neural network analysis of complex traits. Genet. Epidemiol. 14:1101–1106.

Kayaalp E, Treacy E, Waters PJ, Byck, Nowacki P, Scriver C. 1997. Human phenylalanine hydroxylase mutations and hyperphenylalanemia phenotypes: a metanalysis of genotype-phenotype correlations. Am. J. Hum. Genet. 61:1309–1317.

Kruglyak L, Lander ES. 1995a. High resolution genetic mapping of complex traits. Am. J. Hum. Genet. 56: 1212–1223.

Kruglyak L, Lander ES. 1995b. Complete multipoint sib-pair analysis of qualitative and quantitative traits. Am. J. Hum. Genet. 57:439–454.

Lee JH, Reed DR, Li WD, Xu W, Joo EJ, Kilker RL, Nanthakumar E, North M, Sakul H, Bell C, Price RA. 1999. Genome scan for human obesity and linkage to markers in 20q13. Am J Hum Genet. 64:196–209

Li, W-H, Sadler LA. 1991. Low nucleotide diversity in man. Genetics 129:513–523.

Lynch M, Walsh B. 1998. Genetics and analysis of quantitative traits. Sunderland, MA: Sinauer Associates.

MacLean CJ, Bishop DT, Sherman SL, Diehl SR. 1993. Distribution of LOD scores under uncertain mode of inheritance. Am. J. Hum. Genet. 52:354–361.

Mahaney MC, Blangero J, Comuzzie AG, Rainwater DL, VandeBerg, JL Stern MP, MacCluer JW, Hixson JE. 1995. A major locus influences plasma high density lipoprotein cholesterol levels in the San Antonio Family Heart Study: Segregation and linkage analyses. Arterioscler. Thromb. Vasc. Biol. 15:1730–1739.

Mahaney MC, Morin P, Rodriguez LA, Newman DE, Rogers J. 1997. A quantitative trait locus on chromosome 11 may influence bone mineral density at several sites: Linkage analyses in pedigreed baboons. J. Bone Mineral Res. 12(Suppl. 1):S118.

Mahaney MC, Rainwater DL, Rogers J, Cox LA, Blangero J, Almasy L, VandeBerg JL and Hixson JE. 1998. A genome search in pedigreed baboons detects a locus mapping to human chromosome 18q that influences variation in serum levels of HDL and its subfractions. Circulation 98(17)SupplI:5.

Mellersh CS, Langston AA, Acland GM, Fleming MA, Ray K, Wiegand NA, Francisco LV, Gibbs M, Aguirre GD, Ostrander EA. 1997. A linkage map of the canine genome. Genomics 46:326–336.

Mitchell BD, Blangero J, Comuzzie AG, Almasy L, Shuldiner AR, Silver K, Stern MP, MacCluer JW, Hixson JE. 1998. A paired sibling analysis of the beta-3 adrenergic receptor and obesity in Mexican Americans. J. Clin. Invest. 101:584–587.

Mitchell BD, Ghosh S, Schneider JL, Birznieks G, Blangero J. 1997. Power of variance component linkage analysis to detect epistasis. Genet. Epidemiol. 14:1017–1022.

Mitchell BD, Cole SA, Comuzzie AG, Almasy L, Blangero J, MacCluer JW, Hixson JE. 1999. A quantitative trait locus influencing BMI maps to the region of the beta-3 adrenergic receptor. Diabetes 48:1863–1867.

Morin PA, Kanthaswamy S, Smith DG. 1997. Simple sequence repeat (SSR) polymorphisms for colony management and population genetics in rhesus macaques (*Macaca mulatta*). Am. J. Primatol. 42:199–213.

Morin PA, Mahboubi P, Wedel S, and Rogers J. 1998. Rapid screening and comparison of human microsatellite markers in baboons: allele size is conserved, but allele number is not. Genomics 53:12–20.

Morin PA, Wallis J, Moore JJ, Woodruff DS. 1994. Paternity exclusion in a community of wild chimpanzees using hypervariable simple sequence repeats. Molec. Ecol. 3:469–477.

Newman TK, Fairbanks LA, Pollack DB, McGuire MT, Raleigh MJ, Rogers J. 1998. Microsatellite polymorphisms in a captive colony of vervets (*Cercopithecus aethiops sabaeus*). Am. J. Primatol. 45:196–197.

Norman RA, Tataranni PA, Pratley R, Thompson DB, Hanson RL, Prochazka M, Baier L, Ehm MG, Sakul H, Foroud T, Garvey WT, Burns D, Knowler WC, Bennett PH, Bogardus C, Ravussin E. 1998. Autosomal genomic scan for loci linked to obesity and energy metabolism in Pima Indians. Am. J. Hum. Genet. 62:659–68.

Norman RA, Thompson DB, Foroud T, Garvey WT, Bennett PH, Bogardus C, Ravussin E. 1997. Genome-wide search for genes influencing percent body fat in Pima Indians: suggestive linkage at chromosome 11q21-q22. Am. J. Hum. Genet. 60:166–73

Ott J. 1991. Analysis of human genetic linkage, rev. ed. Baltimore: Johns Hopkins University Press.

Pennisi E. 1998. A closer look at SNPs suggests difficulties. Science 81:1787–1789.

Perelygin AA, Kammerer CM, Stowell NC, Rogers J. 1996. Conservation of human chromosome 18 in baboons (*Papio hamadryas*): a linkage map of eight human microsatellites. Cytogenet. Cell Genet. 75:207–209.

Pratley RE, Thompson DB, Prochazka M, Baier L, Mott D, Ravussin E, Sakul H, Ehm MG, Burns DK, Foroud T, Garvey WT, Hanson RL, Knowler WC, Bennett PH, Bogardus C. 1998. An autosomal genomic scan for loci linked to prediabetic phenotypes in Pima Indians. J. Clin. Invest. 101:1757–1764.

Pugh EW, Jaquish CE, Sorant AJM, Doetsch JP, Bailey-Wilson JE, Wilson AF. 1997. Comparison of sib-pair and variance-components methods for genomic screening. Genet. Epidemiol. 14:867–872.

Rainwater DL, Almasy L, Blangero J, Cole SA, Vange-Berg JL, MacCluer JW, Hixson JE. 1999. A genome search identifies major quantitative trait loci on human chromosomes 3 and 4 that influence cholesterol concentrations in small LDL particles. Arterioscl. Thromb. Vasc. Biol. 19:777–783.

Ramsey G. 1998. DNA chips: State-of-the-art. Nature Biotech. 16:40–44.

Rankinen T, Perusse L, Deriaz O, Chagnon M, Nadeau A, Bouchard C. 1999. Linkage of the Na,K-ATPase alpha 2 and beta 1 genes with resting and exercise heart rate and blood pressure: Cross-sectional and longitudinal observations from the Quebec Family Study. J. Hypertens. 17:333–349.

Raskin S, Phillips JA, Kaplan G, McClure M, Vnencak-Jones C, Rozov T, Cardieri JM, Marostica P, Abreu F, Giugliani R, Reis R, Rosario NA, Ludwig N, Pereira L, Faucz F, Gabardo J, Culpi L. 1999. Geographic heterogeneity of 4 common worldwide cystic fibrosis non-DF508 mutations in Brazil. Hum. Biol. 71:111–121.

Reich T, Edenberg HJ, Goate A, Williams JT, Rice JP, Van Eerdewegh P, Foroud T, Hesselbrock V, Schuckit MA, Bucholz K, Porjesz B, Li TK, Conneally PM, Nurnberger JI Jr., Tischfield JA, Crowe RR, Cloninger CR, Wu W, Shears S, Carr K, Crose C, Willig C, Begleiter H. 1998. Genome-wide search for genes affecting the risk for alcohol dependence. Am J Med Genet 81:207–215.

Risch N. 1990. Linkage strategies for genetically complex traits. II. The power of affected relative pairs. Am. J. Hum. Genet. 46:229–241.

Risch N, Giuffra L. 1992. Model misspecification and multipoint linkage analysis. Hum. Hered. 42:77–92.

Roche AF. 1992. Growth, maturation, and body composition: The Fels Longitudinal Study 1929–1991. Cambridge: Cambridge University Press.

Rogers J, Witte SM, Kammerer CM, Hixson JE, MacCluer JW. 1995. Linkage mapping in *Papio* baboons: conservation of a syntenic group of six markers on human chromosome 1. Genomics 28:251–254.

Rotimi C, Comuzzie AG, Lowe WL, Luke A, Blangero J, Cooper RS. 1999. The QTL on chromosome 2 for serum leptin levels is confirmed in African Americans. Diabetes (in press).

Schafer AJ, Hawkins JR. 1998. DNA variation and the future of human genetics. Nature Biotech. 16:33–39.

Schellenberg GD, Kamino K, Bryant EM, Moore D, Bird TD. 1992. Genetic heterogeneity, Down syndrome, and Alzheimer disease. Prog. Clin. Biol. Res. 379:215–226.

Self SG, Liang K-Y. 1987. Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. J. Am. Stat. Assoc. 82:605–610.

Suarez BK, Rice J, Reich T. 1978. The generalized sib pair IBD distribution: its use in the detection of linkage. Ann. Hum. Genet. 42:87–94.

Terwilliger JD, Weiss KM. 1998. Linkage disequilibrium mapping of complex disease: fantasy or reality? Curr. Opin. Biotechnol. 9:578–594.

Towne B, Siervogel RM, Blangero J. 1997. Effects of genotype-by-sex interaction on quantitative trait linkage analysis. Genet. Epidemiol. 14:1053–1058.

Weber JL, May PE. 1989. Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am. J. Hum. Genet. 44: 388–396.

Whittemore AS, Halpern J. 1994. Probability of gene identity by descent: computation and applications. Biometrics 50:109–117.

Wickelgren I. 1999. Discovery of "gay gene" questioned. Science 284:571.

Wijsman EM, Amos CI. 1997. Genetic analysis of simulated oligogenic traits in nuclear families and extended pedigrees: summary of GAW10 contributions. Genet. Epidemiol. 14:719–735.

Williams JT, Blangero J. 1999a. Comparison of variance components and sibpair-based approaches to quantitative trait linkage analysis in unselected samples. Genet. Epidemiol. 16:113–124.

Williams JT, Blangero J. 1999b. Asymptotic power of likelihood ratio tests for detecting quantitative trait loci using the COGA data. Genet. Epidemiol. (in press).

Williams JT, Duggirala R, Blangero J. 1997. Statistical properties of a variance-components method for quantitative trait linkage analysis in nuclear families and extended pedigrees. Genet. Epidemiol. 14:1065–1070.

Williams-Blangero S, Subedi J, Upadhayay RP, Manral DB, Khadka K, Jirel S, Robinson ES, Blangero J. 1998. Attitudes towards helminthic infection in the Jirel population of eastern Nepal. Soc. Sci. Med. 47:371–379.

Williams-Blangero S, VandeBerg JL, Blangero J, Texeira AR. 1997. Genetic epidemiology of seropositivity for *Trypanosoma cruzi* infection in rural Goias, Brazil. Am. J. Trop. Med. Hyg. 57:538–543.

Witte SM, Rogers J. 1999. Microsatellite polymorphisms in Bolivian squirrel monkeys (*Saimiri boliviensis*). Am. J. Primatol. 47:75–84.

Woods SC, Seeley RJ, Porte D, Schwartz MW. 1998. Signals that regulate food intake and energy homeostasis. Science 280:1378–1383.